APPENDIX F

The Use of Phonological Information in Automatic Name Searching

Richard Lutz, Ph.D.

AIPA97

March 25, 1997

PROPRIETARY INFORMATION

2214 Rock Hill Road—Herndon, VA—20170

AIPA97 Paper Presentation:

The Use of Phonological Information in Automatic Name Searching

Richard Lutz, Ph.D.

© *Language Analysis Systems, Inc.*
at the Center for Innovative Technology
2214 Rock Hill Road -- Herndon, VA 20170

Application Area: Automated Data Understanding

Abstract

This paper describes a two-year research effort to incorporate phonological information into automated name searching. Specifically, names represented by standard roman characters are automatically converted to multiple phonetic representations, based on sets of regular expressions that relate character strings to predictable sounds or sound sequences using a widely accepted phonetic notation system, the International Phonetic Alphabet. Names are retrieved when there is an intersection of the regular expression of the query name with regular expressions of names in a preprocessed database. Additional similar names can be retrieved based on the articulatory characteristics of the sound segments contained in the query and database names.

1.0 Introduction

Variation in the spellings of names is a persistent issue in the area of automated name searching in large databases (Hermansen, 1985). In general, the source of spelling variation of names can be analyzed and explained a posteriori. Predicting any individual spelling, however, remains problematic. Sources for spelling variation include: keyboard-based data entry errors (e.g., hitting the wrong key: *Genning* for *Henning*), syntactic variation (e.g., out-of-sequence given name and surname such as *Richard Thomas* for *Thomas Richard*), morphological variation (e.g., truncated strings such as *Rich* or *R* for *Richard*) and semantically-based variation (e.g., nativizations such as *Goldwater* for *Goldwasser*). Of interest in the current paper is variation due to orthographic conventions (e.g., English can represent the same sound in more than one way, as in *Stephen* ~ *Steven*) and articulatory variation (e.g., the *p* in *Thompson* is a predictable spelling of *Thomson* based on principles of articulation). While there are multiple sources of name variation, this paper will present evidence 1) that the inherent ambiguity in the English use of roman characters can be mitigated by multiple mappings to unambiguous phonetic characters and 2) that phonologically-similar names can be retrieved through the analysis of sounds into their articulatory features (i.e., place and manner of articulation). It is based on research conducted from September of 1995 through the present.

2.0 Statement of Problem

Character-based name searching relies on spelling as the basis for calculating distance between the query name and the database name. While spelling using roman characters is not unrelated to pronunciation, the relationship between the two is often inconsistent (Cummings 1988), and the orthographic information (i.e., conventions of the spelling system of a language) is at times misleading. Thus, one spelling may map to multiple pronunciations: *Lutz* can be pronounced to rhyme with *puts*, *cuts* or *shoots*, and at least several additional non-English pronunciations are possible. The converse, of course, is also the case: there may be a number of ways of representing a single pronunciation: *Lewis* and *Louis*, for example, are usually pronounced identically by English speakers.

Character-matching techniques assume a reliable relationship between the orthographic system and the pronunciation. This assumption is flawed because the *goodness of fit* between orthography and pronunciation, especially for English, is *many-to-many*, that is, a given roman character can stand for more than one sound, and an individual sound may be represented in more than one way in the spelling system. Thus, the sound [f][1] can be written as *f* (*Frank*), *ff* (*Taffy*), *ph* (*Phillip*) or even *gh* (*Rough*). Conversely, the *gh* digraph may represent the [f] sound of *Rough*, be silent (*Dough*), or represent [k] (in some pronunciations of *McClaughlin*), [h] (in *Monagham*), [g] (in *McGhee*) or [gh] (across syllable breaks, as in *Bighouse*).

While much name variation can be traced to non-phonological issues, including syntax (order of name segments), aliases (*John Doe* for *John Dillinger*), morphological issues (*Peg* for *Margaret*) or data entry errors, many name variants can be traced to the relationship between orthography and pronunciation. Orally transmitted names, for instance, are especially prone to guesses on the part of the transcriber as to the "official" (i.e., legal) spelling of an individual's name. Language contact can account for some spelling variants as well (French *Beauchamp* and Anglicized *Beecham*), as can transcription from non-roman character sets (*Wachmi* and *Ouakhmi*, *Xie*, *Hsieh* and *Sye*) and sound change over time (e.g., *Leigh* is now pronounced the same as *Lee*).

Additionally, regular (i.e., predictable) processes of speech produce variability in how a name may be written. Thus, the presence of the letter *p* in *Thompson* is an artifact of poor articulatory timing as the articulators move from a nasal [m] to an oral [s]. (The variant spelling *Thomson* reflects a more etymologically justified spelling.)

3.0 Name Representation: Spelling

LAS has been investigating the feasibility and utility of incorporating information about the pronunciation of characters into the automated name searching process. The researchers considered a number of options, including an acoustic-level of representation and character-based rules, and determined that searching of character-based databases could be enhanced to include predictable language-based information about character-to-

[1] Square brackets indicate that a sound is being represented, rather than a spelling.

sound mappings. Specifically, LAS recommended the use of the stock of phonetic symbols known as the International Phonetic Alphabet (IPA), widely used by linguists to represent the inventory of sounds used in the world's languages, and officially adopted by the International Phonetic Association (Laver, 1994). The IPA uses a closed set of symbols to transcribe speech in ways that are interpretable unambiguously by linguists, regardless of the language being described. (See Appendix A.) For example, the symbol [↓] (placed between brackets to indicate that it represents a sound rather than a letter) always stands for a *voiceless* labiodental fricative, as in English ***thigh***, while [] always stands for the equivalent *voiced* labiodental fricative, as in English ***thy***. Thus, IPA disambiguates the English orthographic pattern of using ***th*** to stand for either sound: *thigh* [↓aj] versus *thy* [aj]. A name such as *Gaither*, of course, might be pronounced with either of these sounds, and would thus have two IPA representations, one for each pronunciation: [ge↓r] versus [ger]. There is international agreement by members of the International Phonetic Association, founded in 1889, as to the interpretation of IPA symbols. A re-evaluation of the stock of symbols and special diacritic marks took place at the 1989 IPA Convention in Kiel, and the efforts of the Association have resulted in the unambiguous mapping of sounds onto IPA symbols that transcends individual speakers or languages (Laver, ibid.).

4.0 Mapping Spelling to Sound

The issue of how to predict pronunciation of names from orthography is far from trivial. Two key considerations include that:

- pronunciations of proper names are far less uniform than pronunciations of other vocabulary. The pronunciation of the noun *dough* is more-or-less fixed in English, despite the fossilized spelling that can be traced to an earlier pronunciation. The pronunciation of the name *Lough* is far less certain: individuals named *Lough* may well vary in their pronunciation of the family name and, even if all families named *Lough* could reach a consensus, there is no assurance that those unfamiliar with their consensus would guess that pronunciation. Additionally, some names retain old spellings that map to modern pronunciations in highly improbable ways (e.g., British *Cholmondeley* is commonly pronounced the same as *Chumley*). Claims of "correct" pronunciations carry little weight in terms of name searching;
  and:
- orthographies are language-specific. The pronunciation of the letter *x* regularly maps to [ks] and [z] in English (*Alexander*, *Xenia*), is regularly silent word-finally in French orthography (*LaCroix*), stands for the velar fricative [x], or [s] in Spanish (*México*, *Xochimilco*), and a [dz] or [] in Albanian (*Hoxha*). Additionally, standardized transcription systems from non-roman systems to roman exploit the letter **x** to stand for other, non-English sounds (e.g., Chinese *Xie*, Greek *Xristos*). Finally, any name may be nativized to fit the "borrower" language: spellings of non-Anglo names may be pronounced according to English orthographic conventions (e.g., French *Duquesne* pronounced [dukwzni].)

5.0 Writing IPA Conversion Rules

IPA is an effective notational system for representing pronunciation. LAS has written sets of rules that relate spellings to sounds. The rules are language-based, with sets of rules operating for Arabic, Mandarin Chinese, Hispanic and Anglo names. The rules assume:

- 26-character sets of roman letters, absent all diacritic markings, including accent marks or tone indicators;
- English speakers, either naïve or expert in the language of origin;
- one spelling can map to multiple pronunciations.

The rule sets were written to specific development databases made of single name elements, either surname or given, and taken from a variety of sources, including the U.S. Census list of the most frequent names in the U.S. and large U.S. databases of names from other countries. The names were manually tagged as "Arabic", "Mandarin Chinese", "Hispanic" and "Anglo", where "Anglo" was loosely interpreted to include Western European Germanic names (including Dutch and German). A team of linguists used a variety of sources to determine possible pronunciations, including native speaker knowledge and textual information (e.g., Cummings, 1988, Hanks and Hodges, 1989, 1990, Symonds, 1986). In general, rules were written broadly in order to ensure that most plausible pronunciations were captured. The Arabic and Mandarin Chinese rules included transcription variation (e.g., Chinese pinyin, Wade-Giles and Yale conventions of rendering Chinese names into roman script, as in *Xie/Hsieh/Sye*). The sample Anglo rule below is interpreted to mean that the letters *sc* preceded by anything and followed by the letters *le* can be pronounced as [s] or [sk] (e.g., *Muscle* and *Mosclin*):

sc/ anything ___ *le* → [sk?]

Rules were implemented using standard regular expression notation. The following table shows a sample query and the names returned from a data file containing the 88,799 most frequent surnames from the U.S. census:

Search on SMITH
SMITH
SMYTH
SMITHE
SMIT
SMYTHE
SMIDT
SMIHT
SZMIDT

Figure 1 Search on name *SMITH*

As an example of the advantages of matching on IPA, consider a query on the name *Lee*. Converted to the IPA string [li], exact matches with numerous spelling variants are automatic, including *Leigh* and *Li*. Typical character-based matches will fail to retrieve *Leigh* or *Li*, since the percentage of character overlap is minimal. Conversely, a standard index matching system such as Soundex will categorize *Lee* and *Li* identically, but will still miss *Leigh*, given the presence of a salient letter (*g*), and will retrieve a large number of names of low relevance, including *Lu*, *Liao*, *Low*, *Louie*, *Lahoya and Lehew*.

6.0 Phonological Processes

In addition to predictable spelling variation, rules were written to account for predictable articulatory processes (MacKay, 1987; Wolfram and Johnson, 1982). For example, the variant spellings of *Thomson ~ Thompson, Simson ~ Simpson, Demsey ~ Dempsey, etc.* can be accounted for by regular movement of the velum (i.e., the soft palate) from a bilabial nasal [m] to an oral [s]. Production of an intrusive bilabial oral [p] is entirely a result of the timing of the movement from nasal to oral articulation. LAS incorporated likely articulatory variation into the IPA rule sets. Thus, a query of the name *Thomson* will retrieve the variant *Thompson* as an exact match.

7.0 Testing the Rule Sets

To test the net effect of the Orthography-to-IPA rules, LAS conducted a controlled test of the rules by randomly selecting 157 test names from a database of 55,545. The database contained names that were from sources identified as Arabic, Mandarin Chinese, Hispanic and Anglo (again, broadly defined). A native speaker of educated standard American English was asked to record the 157 test names using pronunciations of his choosing. The audio recordings were played for native speakers of American English, who were asked to write one or more "likely" spelling for each name. LAS elicited 3,689 variants in all by playing the recordings to native speakers of American English. The variant spellings were then used as test query names to calculate the retrieval rates of the original name spellings. Overall, 69% of all variant spellings were retrieved by the IPA rules. However, qualitative analysis of the results showed that approximately 23% of the variant names not retrieved were due to perceptual mishearings of the recorded names. For example, the variant spellings of the test name *Baughn* predictably included *Bahn*, *Baun*, and *Bonn*, and the IPA Conversion Rules succeeded in mapping all to the original test name spelling. However, a fourth elicited spelling, *Vaughn*, was not predicted, and the IPA Conversion Rules did not map it to *Baughn*. The mishearing of [v] for [b] is not unusual, given the acoustics shared by the two sounds. The IPA Conversion Rules, which include regular <u>articulatory</u> variants such as *Thomson/Thompson*, were purposely not intended to retrieve <u>perceptually</u> similar names during the current phase of research.

8.0 Fuzzy Matches: Articulatory Similarity

At the heart of the research has been an effort to improve the automatic name searching process by retrieving names that are *similar* to the query name. The IPA Conversion Rules are able to capture a good deal of name variation that can be attributed to orthographic sources, whether intralingual (e.g., *Leigh/Lee*) or interlingual (e.g., transcriptions to roman orthography from Chinese: *Xie* ~ *Hsieh* ~ *Sye*). An additional goal has been to retrieve names that are not phonologically identical to the query name, but that a careful analyst would like to consider before abandoning a search. Thus, while spelling variants of the name *Benke* include *Behnke* and *Benck*, the analyst might want to consider names that seem phonologically close to the query name without being a predictable variant (e.g., *Benge*, *Bankey* and perhaps even names like *Penke*, *Panke* or *Bentsche*). While most search algorithms permit fuzzy matches, these are invariably based on calculations of number of characters shared. From the perspective of character matching, the letter *b* is as different from the *p* as it is from *x*, *y* or *z*. Thus, to permit retrieval of *Penke* for *Benke* is to require retrieval of any name that differs from the query by the first character, including *Xenke*, *Yenke* and *Zenke*. This clearly does not follow any phonologically reliable principle, and significantly reduces the efficiency of automatic retrieval. Even indexed systems, such as Soundex, group letters as either co-indexed or unrelated. Thus, while Soundex is often called "phonetic" because it groups letters that share some phonological characteristics, it cannot compare the degree to which two sounds, or indeed two names are related: it lacks granularity. Thus, Soundex would treat *Benke*, *Penke* and *Panke* as identical rather than similar. Soundex would exclude *Bentsche* from the group because of the letter *t* in the spelling, in effect treating *Bentsche* as being equally distant from *Benke* as from *Smith*.

It is clear, however, that sound segments can be analyzed in terms of their articulatory characteristics, and that some sounds fall into natural categories, such as vowels and consonants. Properties of sounds have been described in detail by a number of linguistic analyses according to place and manner of articulation (e.g., [p] and [b] are both articulated at the lips by complete blockage of the air flow and sudden release of pressure). One of the best known descriptions of phonetic classification is that of the American linguists Chomsky and Halle (1968). All the distinct sounds of American English can be described using 15 distinctive features (see Appendices B and C). By classifying sounds according to these distinctive features, a fairly clear picture emerges of how close any two sounds are to one another. Thus, [p] and [b] differ by just one feature, voicing, while [p] and [f] differ by three and [p] and [v] by four. In general, articulatory distance can be counted in terms of how many articulatory characteristics sounds share.

LAS created a file of feature differences between pairs of sounds, essentially mapping phonetic features onto IPA notation. By relaxing the threshold of allowable differences, increasingly distant sounds are retrieved. Thus, by permitting matches of IPA characters that are not exact matches, names are retrieved that are phonologically close. Even IPA sound-to-sound comparisons yield interesting sets of names for comparison. By relaxing

retrievals to include single feature differences, a search of the name *Smith* now brings back these additional names:

Search on *SMITH*
Feature Difference Threshold: 1

SMID
SMEAD
SNITH
SNIPE
SNIDE
SNEED
SNEAD
SNAPE
SNEATH

Figure 2 Fuzzy Search on *Smith* measuring Phonetic Feature Differences

Viewed in physiological terms, this is reasonable. Phonetic features refer to salient characteristics of articulation, so that differences generally reflect how likely it is that any two sounds would be articulated in place of another. There are numerous additional factors, of course, that ought to be considered in measuring how similar two names are to one another articulatorily.

9.0 Final Sorting of Names Retrieved

The names retrieved by searches on phonetic features may not all be of equal relevance to the query name. Additional factors are under consideration to sort names retrieved, based on a variety of phonological characteristics.

9.1 Sonority Level

The differences in phonetic features generally express the amount of effort needed to move articulators from one sound to another. The sounds [p], [t] and [k] form a natural class of voiceless stop consonants — identical in manner of articulation. All are extremely common in the world's languages, and are among the first acquired by children. They differ in place of articulation, and this is reflected in feature differences. However, manner of articulation is probably a better measure of energy expenditure than is place of articulation: voiceless stops are all extremely low in sonority, that is, the amount of energy needed to produce a sound. Vowels, on the other hand, require much more effort: they, in essence carry the sound wave. In order for feature differences to effectively measure level of effort required, differences should be weighted according to sonority level. In general terms, sounds fall into nine levels of sonority, with voiceless stops [p], [t] and [k] at the

low end and the vowels [ ] as in *father* and [⊕] as in *fan* at the most sonorous end (Ladefoged, 1982). Sorts of names retrieved ought to consider the sonority value of sounds. This might be accomplished by weighting phonetic features or by a more complicated comparison of sonority level contours of names or syllables.

9.2 Syllabification

Additionally, in languages that time segments based in part on stress patterns, it is reasonable to compare stressed syllables to one another. In the following example, names have been aligned in terms of substrings, in this case corresponding to syllables:

Chester:	[τ♣ στρ]
Chesterton:	[τ♣ στρ τν]
Winchester:	[ω ν τ♣ στρ]

Both in terms of articulatory effort (sonority) and psychological salience, it would be misleading to treat all three occurrences of the substring [t♣] as equivalent: stress clearly must be included in the equation. LAS has written a syllabifier that automatically parses English IPA strings, including names, according to a set of rules. Future research will investigate the possibility of ranking similar names through analysis at the syllabic level. Syllabic level analysis has the strength of lining up comparable substructures of names. All syllables share the same internal structures (i.e., onset of the syllable, nucleus, and coda), and alignment by syllable enables meaningful comparisons of internal structures of names (where a period represents the syllable break):

Linda	[λ ν . δ]
Lisa	[λ ι _ . σ]

Note that in the above example, the coda (i.e., end) of the first syllable in *Linda* is filled by [n] but empty in *Lisa*, as indicated by the underscore. A meaningful comparison of the two names would compare the [n] of *Linda* to an empty coda rather than to the [s] in the onset (i.e., beginning) of the second syllable of *Lisa*.

9.3 Position in Name

Some weight ought to be given to absolute initial position in names. Many indexed systems, including Soundex, key names to the initial letter. This is, of course, problematic, since the initial letter may be silent or part of a digraph (e.g., *Knox*, *Philip*). However, indexing on the first sound, or at least considering the first sound as more significant than sounds in other positions may be warranted. This, like syllable-level comparisons, will probably be a factor in final sorting of names retrieved.

9.4 Non-Phonological Factors in Sorting of Names Retrieved

Certainly, it must be acknowledged that non-phonological levels of analysis may be critical to any useful definition of *similarity*. Morphological units – word parts that may contain

semantic information, including prefixes and suffixes — such as *Mc-*, *-ton*, and *-sky* are likely sources of variations. Thus, *Lubin* and *Lubinsky* are critically related (in terms of their roots), while *Lubin*, *Rubin* and *Lupine* are very close in terms of articulation. The morphological factor could be handled efficiently with a look-up list of morphological elements, but this remains outside the current scope of this project.

Similarly, orthography itself might play a useful role in the final sort of names retrieved. The following names retrieved for a fuzzy search on the name *Bucket* have been sorted using a simple sort on letters.

Search on **BUCKET**
Feature Difference Threshold: 1

BYXKETT
BEXKET
BIXKET
BYΘYET
BEXKETT
BIXKETT
BYXHHEIT
BOΘYET
BAΘYET
BOXHAT
BYXHITE
BEXKΩITH
BAXOT
BOOKOYT
BAXOTE
BOYΘYET
BEKHIT
BOΩXYTT
BEΘYETTE

Figure 3 Search on *Bucket* Sorted by Spelling

Current plans are for a final ranking of names retrieved based on a combination of factors, including number of syllables, stress, weighting of features by sonority levels and name-initial segments.

10.0 Conclusions

In sum, automatic name searching can benefit in three ways from incorporation of phonological information:

- leveling differences due exclusively to orthographic mapping;
- leveling differences due to predictable phonological processes, such as intrusive consonants; and
- retrieving additional names that contain phonologically similar sounds to those of the query name.

Having retrieved phonologically relevant names, a phonologically-enhanced name search engine can then sort names using a multiple factor weighting scheme.

LAS views this technology as extremely promising, offering a tool to enhance current automatic name searching, increasing chances of retrieving name variants that character-based systems miss by retrieving and sorting names in a phonologically principled way.

Appendix A: Descriptions of IPA Symbols

Phonetic symbol	Description	Example
p	voiceless bilabial stop	p in the English name Peter
b	voiced bilabial stop	b in the English name Buddy
Φ	voiceless bilabial fricative	f in the Japanese name Fujimori
[illegible]	voiced bilabial fricative	b in the Spanish word saber
m	bilabial nasal	m in the English name Mary
←	voiced rounded palatal approximant	u in the French name Nuit
f	voiceless labio-dental fricative	f in the English name Fred
v	voiced labio-dental fricative	v in the English name Vera
[illegible]	voiced labio-dental nasal	n in the Italian word anfora
t	voiceless alveolar stop	t in the English name Ted
d	voiced alveolar stop	d in the English name Doug
θ	voiceless apico-dental fricative	th in the English name Theodore
x	voiced apico-dental fricative	th in the English name Rather
s	voiceless alveolar fricative	s in the English name Sam
z	voiced alveolar fricative	z in the English name Zachary
n	voiced alveolar nasal	n in the English name Nathan
l	voiced alveolar lateral	l in the English name Linda
◷	voiceless alveolar lateral fricative	ll in the Welsh name Llewellyn
☹	voiced alveolar lateral fricative	dhl in the Zulu word dhla (to eat)
☐	voiced alveolar continuant	r in the English name Richard
r	voiced apico-alveolar trill	r in the Spanish name Ricardo
☼	voiced alveolar flap	tt in the English name Ritter
[illegible]	voiceless retroflex stop	as in the Arabic name Tariq
⬇	voiced retroflex stop	as in the Arabic word difda' (frog)
▪	voiceless retroflex fricative	as in the Arabic name Sabir
◷	voiced retroflex fricative	as in the Arabic name Dhafir
✎	voiced retroflex nasal	Marathi (India)
⇧	voiced retroflex lateral approximant	Marathi (India)
"	voiced retroflex flap	d as in Hindi dal (lentil stew)
♦	voiceless palato-alvoelar fricative	sh in the English name Sheila
C·	voiced palato-alveolar fricative	z in the English word azure
☑	voiceless alveo-palatal fricative	x as in the Chinese name Xia
✓	voiced alveo-palatal fricative	ï in the Polish word ïle
t♦	voiceless palato-alvoelar affricate	ch in the English name Charlie
dC·	voiced palato-alveolar affricate	j in the English name Jennifer
[illegible]	voiced palatal nasal	■➀ in the Spanish word Do■➀a
◈	voiced palatal lateral approximant	ll in the Spanish word calle (street)
k	voiceless velar stop	k in the English name Kim
g	voiced velar stop	g in the English name Gary
x	voiceless velar fricative	x in the Spanish name Jose
[illegible]	voiced velar fricative	g in the Spanish word luego (later)
ŏ	voiced velar nasal	ng in the English name Bing

Appendix A: Descriptions of IPA Symbols (Continued)

Phonetic symbol	Description	Example
[illegible]	voiceless velar lateral	l in the Polish Walesa
⇖	voiceless labio-velar approximant	wh as in the English name White (for some speakers)
w	voiced bilabial approximant	w in the English name Wayne
q	voiceless uvular stop	as in the Arabic name Qasim
G	voiced uvular stop	Eskimo and Tehrani Persian
[illegible]	voiceless uvular fricative	ch as in the German word Buch
[illegible]	voiced uvular fricative	r in some Parisian pronunciations of the French name RenJe
N	voiced uvular nasal	n in the Eskimo word eNima (melody)
R	voiced uvular trill	r in the French name RenJe
⇨	voiceless pharyngeal fricative	h as in the Arabic name Muhammad
⌦	voiced pharyngeal fricative	as in the Arabic name Sa'ad
[illegible]	voiceless glottal stop	tt as in the English name Sutton or the word mitten
h	voiceless glottal fricative	h in the English name Henry
▫	voiced glottal fricative	h as in English between voiced sounds, as in the word manhood
y	high front rounded vowel	u in the French word lune (moon)
●	high central unrounded vowel	as in the Russian word s●n (son)
✱	High central rounded vowel	u as in the Norwegian hus
✪	high back unrounded vowel	u as in the Japanese name Kazu
u	high back rounded vowel	ou as in the French word tout
[illegible]	upper mid-front rounded	ŏ as in the German name Sch'nfeld
➚	upper mid-back unrounded vowel	as in the Shan (Burma)word 'ko (salt)
o	upper mid-back rounded	o as in the English name Mona
I	semi-high front unrounded vowel	y as in the English name Lynn
ε	lower mid-front unrounded	e as in the English name Deborah
❋	lower-mid front rounded vowel	oeu as in the French word oeuf (egg)
▯	lower-mid back unrounded vowel	u as in the English name Tupperman
❷	lower-mid back unrounded	o as in the English name Ford
F	open front unrounded vowel	a as in the English name Hal
❶	open central unrounded vowel	a as in the Portuguese word para (for)
[illegible]	low front unrounded vowel	a as in the French word patte (paw)
α	low central unrounded vowel	â as in the French name Delâtre or the word p>te (paste or dough)
①	low back rounded vowel	o as in the British English word hot
★	mid central unrounded vowel	e & a as in the English name Belinda
U	semi-high back rounded vowel	u as in the English name Butch
e	upper-mid front unrounded	a as in the English name Mable
i	high front unrounded vowel	first e in the English name Pete
[illegible]	rhotacized mid-vowel	ea as in the English name Heather

Appendix A: Descriptions of IPA Symbols (Continued)

Phonetic symbol	Description	Example
t☑	voiceless alveo-palatal affricate	j as in the Chinese name Jin
t☑'	voiceless aspirated alveo-palatal affricate	q as in the Chinese name Qiu
ts	voiceless unaspirated dental affricate	ts as in the Chinese name Tsang
ts'	voiceless aspirated dental affricate	c as in the Chinese name Cao
⑦	bilabial click	as in Southern Bushman languages
|	dental (alveolar) click	as in Bushman
!	palatal click	as in Bushman
⇜	palato-alveolar click	as in Hottentot
④	alveolar lateral click	as in Bushman, Zulu

Appendix B: Description of Phonetic Features

A. Major class features:

1. ***Syllabic***
Forms the central peak of a syllable. Vowels are usually +syllabic, consonants are usually -syllabic, but some (like [l]) may be syllabic (as in "riddle")

2. ***Sonorant***
Minimal constriction in the mouth. Vowels, as well as [n], [m], [r], [l], [w] are all +sonorant. Most other consonants are -sonorant.

3. ***Consonantal***
Obstruction along a central point in the mouth. All English sounds except vowels and glides ([w] and [y]) are +consonantal.

B. Manner of Articulation Features:

4. ***Continuant***
Continued air movement through the mouth during sound production. This feature contrasts fricative sounds like [f] and [v] with non-continuants like [p] and [b].

5. ***Strident***
Narrow obstruction through which air escapes, producing hissing or "white noise". [s], [z], [f], [v] and the sounds in **church** and **judge** are +strident. This is the most acoustically-based feature in this list.

6. ***Delayed Release***
Gradual release of air. In English, it is used to distinguish the sounds in **church** and **judge** from [t] and [d]

7. ***Nasal***
Soft palate at the back of the mouth is lowered and air goes into nose. In English, [n], [m] and [ŋ] (the final sound in **king**) are +nasal.

8. ***Lateral***
Side(s) of tongue lowered so that air escapes along side, as in English [l].

C. Place of articulation:

9. ***Anterior***
Obstruction of mouth anywhere from gum ridge forward to lips. English [p], [b], [m], [f], [v], and [ð] (as in **the**) are all +anterior.

10. ***Coronal***
Front of the tongue raised. The sounds [t] and [d] are +coronal. Sounds like [k] and [g] are -coronal.

11. ***High***
Body of tongue raised. [j] (as in yellow), and the vowel [i] (as in **feet**) are +high.

Appendix B: Description of Phonetic Features (Continued)

12. ***Low***
Body of tongue lowered. The vowels [[illegible]] as in back and [[illegible]] as in father are +low.

13. ***Back***
Body of tongue moved back. The sounds [k] and [g] and the vowel [u] as in boot are +back.

14. ***Tense***
Root of tongue muscle tensed. The vowel [[illegible]] (as in feet) is +tense. The vowel [[illegible]] as in fit is -tense.

15. ***Round***
Lips pursed or rounded. English vowel [u] (as in boot) is +round, while [[illegible]] (as in beet) is -round.

Appendix C: Phonetic Features for [p], [b] and [f]

Phonetic Features	[p]	[b]	[f]
syllabic	-	-	-
sonorant	-	-	-
consonantal	+	+	+
anterior	+	+	+
coronal	-	-	-
high	-	-	-
low	-	-	-
back	-	-	-
continuant	-	-	+
strident	-	-	+
delayed release	-	-	-
voiced	-	+	-
nasal	-	-	-
lateral	-	-	-
round	-	-	-

References

Chomsky, Noam and Halle, Morris. *The Sound Pattern of English*. Harper & Row, New York, 1968.

Cummings, D. W. *American English Spelling*, The Johns Hopkins University Press, London, 1988.

Hanks, Patrick, and Hodges, Flavia. *A Dictionary of First Names*, Oxford University Press, Oxford, 1990.

Hanks, Patrick, and Hodges, Flavia. *A Dictionary of Surnames*, Oxford University Press, Oxford, 1989-90.

Hermansen, John C. *Automatic Name Searching in Large Data Bases of International Names*, unpublished dissertation, Georgetown University, Washington, D.C., 1985.

Ladefoged, Peter. *A Course in Phonetics*, Harcourt Brace Jovanovich, Publishers, San Diego, 1982.

Laver, John. *Principles of Phonetics*, Cambridge University Press, 1994.

MacKay, Ian. *Phonetics: The Science of Speech Production*, Little, Brown, and Company, Boston, 1987.

Symonds, Martin A. *Mandarin Pronunciation*, Taipei Language Institute, Taipei, 1986.

Wolfram, Walt and Johnson, Robert. *Phonological Analysis: Focus on American English*, The Center for Applied Linguistics & Harcourt Brace Jovanovich, Inc., 1982.